

# HASIL CEK\_60020397\_C73- Komparasi Image Matching Menggunakan Metode K-Nearest Neighbor (KNN) dan Metode Support Vector Machine (SVM) *by Imam Riadi 60020397*

---

**Submission date:** 11-Dec-2020 11:07AM (UTC+0700)

**Submission ID:** 1471719661

**File name:** -Nearest\_Neighbor\_KNN\_dan\_Metode\_Support\_Vector\_Machine\_SVM.pdf (989.75K)

**Word count:** 4001

**Character count:** 23525

## Komparasi *Image Matching* Menggunakan Metode *K-Nearest Neighbor (KNN)* dan Metode *Support Vector Machine (SVM)*

<sup>1</sup> Rusydi Umar <sup>1\*</sup>, Imam Riadi <sup>2\*\*</sup>, Dewi Astria Farook <sup>3\*</sup>

<sup>\*\*</sup> Program, Studi Teknik Informatika, Universitas Ahmad Dahlan <sup>1,3</sup>

<sup>\*</sup> Program, Studi Sistem Informasi, Universitas Ahmad Dahlan <sup>2</sup>

[rusydi@mti.uad.ac.id](mailto:rusydi@mti.uad.ac.id)<sup>1</sup>, [imam.riadi@is.uad.ac.id](mailto:imam.riadi@is.uad.ac.id)<sup>2</sup>, [dewi1907048014@webmail.uad.ac.id](mailto:dewi1907048014@webmail.uad.ac.id)<sup>3</sup>

### Article Info

#### Article history:

Received 2020-07-23

Revised 2020-10-10

Accepted 2020-10-22

#### Keyword:

*Image Matching*,  
*Classification*,  
*KNN*,  
*SVM*

### ABSTRACT

*Image matching is the process of finding digital images that have a degree of similarity. Matching images using the classification method. In measuring image matching, the images used are original logo images and manipulated logo images. Comparison of classification algorithms from the two methods namely K-Nearest Neighbor (KNN) and Support Vector Machine with Sequential Minimal Optimization (SMO) optimization used to calculate matches based on accuracy values. The K-Nearest Neighbor (KNN) classification method is based on proximity or K calculations while the Support Vector Machine (SVM) classification method measures the distance between the hyperplane and the nearest data. Image match values are measured by Precision, Recall, F1-Score, and Accuracy. The image matching steps start from the preparation of data processing, extraction of HSV color features and shapes, then the classification stage. Digital images are used as many as 10 images consisting of one original logo and 9 manipulated logos. In the classification testing stage, using the WEKA application by applying the 10-fold cross-validation method. From the results of tests conducted that the closest k-neighbor (KNN) classification method is 80% and has a  $k = 0.889$  which is quite good in measuring proximity, while the SVM classification method is 70%. The results of this image matching comparison can be concluded that the K-Nearest Neighbor classification method works better than SVM for image matching.*



<sup>2</sup> This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

### 1. PENDAHULUAN

*Image Matching* atau disebut pencocokan citra merupakan salah satu metode dari pengolahan citra yang digunakan untuk mencari citra lain yang sejenis atau citra yang memiliki kecocokan atau kesamaan. Pencocokan gambar menjadi salah satu teknik penting dalam pemrosesan gambar dan telah banyak digunakan dalam analisis gambar diantaranya gambar medis, rekonstruksi 3D dan sebagainya[1]. Sebuah citra digital dengan beberapa kesamaan atau kemiripan dapat diukur berdasarkan parameter ukur yang digunakan. Untuk menghitung *image matching* pada citra digital yang memiliki tingkat kesamaan, diukur seberapa mirip citra tersebut yaitu dengan pencocokan citra. Untuk mengatasi permasalahan *plagiarism* pada sebuah karya seni maka metode pencocokan

citra (*Image Matching*) dapat dilakukan pada citra digital tersebut. Komparasi image dilakukan dengan menggunakan metode klasifikasi. Citra yang akan diklasifikasi merupakan *image* logo asli dan *image* logo manipulasi. Logo yang digunakan untuk pencocokan citra adalah logo berformat .png. Dengan menggunakan Algoritma *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)*. Tahapan dalam mengklasifikasi citra digital yaitu persiapan data, ekstraksi fitur ciri warna HSV dan bentuk, pelatihan, pengujian dan pengukuran akurasi. Pengolahan citra dilakukan dengan mengumpulkan citra yang digunakan sebagai citra masukan untuk mendapatkan informasi yang tertera pada setiap citranya.[2] Tingkat pencocokan citra (*image matching*) diketahui berdasarkan nilai *Precision*, *Recall*, *F-Measure* dan Akurasi untuk mengetahui pencocokan citra (*Image*

Matching). Dari komparasi image dengan metode *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) diharapkan dapat menunjukkan tingkat kemiripan citra yang diuji dan dilatih sehingga *plagiarism* pada sebuah karya dapat diketahui dengan tingginya nilai akurasi pada pencocokan citra tersebut.

Pada penelitian terdahulu, metode klasifikasi dapat digunakan dalam bidang apa saja. Diantara yang dilakukan pada penelitian [3] terkait klasifikasi kematangan buah apel manalagi dengan menggunakan algoritma *Naïve Bayes* dan ekstraksi fitur citra digital. Ekstraksi fitur yang digunakan sebagai data klasifikasi yaitu rerata intensitas, standar deviasi, smoothness, skewness, energy dan entropy. Dalam mengetahui kematangan buah, tingkat kematangan buah dibagi menjadi dua yaitu apel matang dan apel mentah. Klasifikasi tersebut menghasilkan nilai akurasi sebesar 63% dengan menggunakan metode algoritma *naïve Bayes*.

Penelitian yang dilakukan [4] yaitu *support vector machine* untuk klasifikasi citra jenis daging berdasarkan tekstur menggunakan ekstraksi ciri *Grey Level Co-Occurrence Matrices (GLCM)* data yang digunakan dalam penelitian ini adalah citra daging kuda, kerbau, sapi dan kambing dengan jarak pengambilan 20 cm, 30 cm dan 40 cm. Penelitian menghasilkan tingkat pengenalan yang cukup tinggi dalam mengenali citra daging yaitu 87,5% dengan jarak pengambilan 20 cm dan jarak piksel tetangga  $d=2$  arah GLCM  $135^\circ$ .

Penelitian [5] yang dilakukan dalam dunia kesehatan, digunakan pada klasifikasi sel darah putih berdasarkan ciri warna dan bentuk dengan metode *K-Nearest Neighbor (KNN)* dalam proses pengujian data yang digunakan sebanyak 100 citra. Dari hasil tersebut pengujian segmentasi menunjukkan akurasi sebesar 78% dan pengujian klasifikasi sebesar 64% cukup baik dalam klasifikasi sel darah putih.

Penelitian [6] melakukan perbandingan metode SVM, RF, dan SGD untuk penentuan Model Klasifikasi Kinerja Programmer pada aktivitas media sosial dengan menggunakan *text preprocessing* untuk melakukan normalisasi data teks, dan seleksi fitur setelah pembobotan kata dengan metode TF-IDF maka hasil dari performa klasifikasi dari ketiga metode tersebut adalah *support vector machine (SVM)* adalah 81,3%, *Random Forest (RF)* 74,4%, dan *Stochastic Gradient Descent (SGD)* 80,1%. Metode SVM terpilih sebagai metode terbaik dalam penentuan klasifikasi kinerja programmer pada aktivitas media sosial.

Pada penelitian [7] Klasifikasi citra makanan untuk mengedukasi jenis makanan yang baik, dengan menggunakan ekstraksi fitur warna Hue Saturation Value (HSV) dan ekstraksi fitur Local Binary Pattern (LBP) proses yang dilakukan pada penelitian ini dimulai dari *pre-processing*, ekstraksi fitur warna HSV dan menggunakan LBP menghasilkan nilai keabuan dalam histogram. Maka hasil ekstraksi fitur dari setiap citra tersebut kemudian digunakan untuk klasifikasi menggunakan *Naïve Bayes*. Hasil dari klasifikasi menggunakan metode HSV adalah 65% dan hasil dari metode LBP sebesar 60%.

Dari beberapa penelitian metode klasifikasi sangat membantu dalam memproses citra digital. Penelitian ini bertujuan untuk mengetahui kinerja dari algoritma dalam mengklasifikasi citra dan mengetahui hasil akurasi yang kemudian di komparasi dari kedua metode yaitu *K-Nearest Neighbor (K-NN)* dan *Support Vector Machine (SVM)* dalam pencocokan citra. Berdasarkan pada nilai *Precession*, *Recall*, *F-Measure* dan Akurasi antara image logo asli dan image logo temanupulasi.

## II. TINJAUAN PUSTAKA

### A. Pengolahan Citra Digital

Citra merupakan sebuah gambaran, kemiripan, atau imitasi dari suatu objek. Citra dapat berupa foto Sinyal video (gambar di televisi) atau bersifat digital yang dapat langsung disimpan dalam media penyimpanan [8]. Citra digital merupakan citra yang diolah oleh komputer. Proses yang dilakukan untuk mendapatkan citra digital sampai pencitraan yaitu dengan akuisisi citra [9].

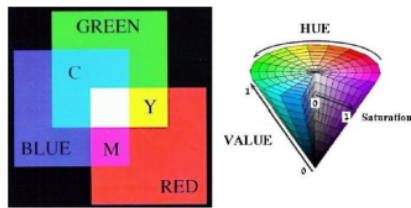
Pengolahan Citra merupakan salah satu area paling penting dari aplikasi multimedia. Citra yang telah dikompresi, didekompresi, dikirim melalui internet, dimodifikasi atau bahkan terdistorsi dalam berbagai citra, ini semua dapat mempengaruhi kualitas dari citra tersebut. Kualitas dari citra inilah yang merupakan elemen kunci bagi multimedia untuk dapat menilai perubahan, kualitas citra, memprediksi dan memperbaikinya.

### B. Ekstraksi Ciri

Ekstraksi ciri merupakan suatu proses yang dapat dilakukan untuk mengambil bermacam ciri yang ada pada sebuah citra [10]. Ekstraksi ciri warna yang dilakukan pada image logo yang telah berhasil disegmentasi. Proses ekstraksi ciri menggunakan ekstraksi ciri warna dan bentuk. Tujuan dari ekstraksi ciri warna dan bentuk adalah untuk mendapatkan nilai data numerik ciri masing-masing dalam parameter ciri citra tersebut. Pada penelitian ini menggunakan parameter ekstrasi ciri Hue, Saturation, Value, Luas Area, Tepi area, dan Kebundaran. Berikut parameter dari ekstraksi ciri citra tersebut:

#### 1. HSV (Hue Saturation Value)

Model HSV didefinisikan warna dalam teknologi Hue, Saturation dan Value. Warna HSV pertama kali diperkenalkan oleh A.R Smith tahun 1978. Hue menyatakan warna sebenarnya seperti warna merah, violet dan kuning untuk warna kemerahan, kehijauan [11]. Saturation merupakan kemurnian atau kekuatan warna. Value adalah kecerahan warna [12]. Nilai pada HSV antara 0-100%, warna hitam bernilai 0 semakin tinggi nilainya akan terlihat warna baru. Seperti pada gambar 1 koordinat warna HSV. [13]



Gambar 1. Koordinat warna HSV

Pada gambar 1 menunjukkan sistem koordinat HSV sebagai hexacone. Dasar hexacone berwarna hitam dengan HSV = (0,0,0). Jadi apabila menggunakan model HSV warna kulit orang afrika maka di sekitar dasar hexacone inilah koordinatnya. Kebanyakan gambar warna dicatat sebagai (R, G, B). Warna didefinisikan oleh (R, G, B) dimana R, G, B dinormalisasi dari 0.0 sampai dengan 1.0, dan setara dengan warna (H, S, V) yang ditentukan oleh suatu set formula. MAX merupakan nilai maksimum dari (R, G, B) dan MIN merupakan nilai minimum dari model tersebut.

$$H = \begin{cases} \frac{G - B}{\text{Max} - \text{Min}} \times 60, & \text{if } R = \text{Max} \\ 120 + \frac{B - R}{\text{Max} - \text{Min}} \times 60, & \text{if } G = \text{Max} \\ 240 + \frac{R - G}{\text{Max} - \text{Min}} \times 60, & \text{if } B = \text{Max} \end{cases} \quad (1)$$

$$S = \frac{\text{Max} - \text{Min}}{\text{Max}}; V = \text{Max} \quad (2)$$

Dimana H bervariasi dari 0 sampai dengan 360, yang menunjukkan sudut dalam derajat, S dan V bervariasi dari 0.0 sampai dengan 1.0. Untuk memudahkan pemahaman tentang metode ini dengan menggambarkan bagaimana Hue dihitung. Tingkat dari Hue ditentukan dengan terlebih dahulu mencari nilai dominan diantara (R, G, B) dan menghitung nilai Hue berdasarkan langkah-langkah berikut. Jika warna dominan adalah merah maka:

$$\text{Hue} = \frac{G - B}{\text{RGB}_{\text{Max}} - \text{RGB}_{\text{Min}}} \times 60 \quad (3)$$

Jika warna dominan adalah warna hijau kemudian Hue dihitung sebagai:

$$\text{Hue} = 120 + \frac{B - R}{\text{RGB}_{\text{Max}} - \text{RGB}_{\text{Min}}} \times 60 \quad (4)$$

Demikian pula dengan warna yang dominan biru:

$$\text{Hue} = 240 + \frac{R - G}{\text{RGB}_{\text{Max}} - \text{RGB}_{\text{Min}}} \times 60 \quad (5)$$

Pertimbangan ini merupakan ringkasan dari persamaan (1).

## 2. Area

Area pada citra dapat diperoleh dengan menghitung jumlah piksel bernilai 1. Luas area citra didapat dari ukuran *image* logo asli dan *image* logo termanipulasi.

## 3. Perimeter (Tepi Area)

Perimeter atau tepi area dapat diperoleh menghitung piksel terluar dari objek. Dengan demikian akan mendapat hasil tepi area yang kemudian akan dihitung jumlah pikselnya.

## 4. Eccentricity

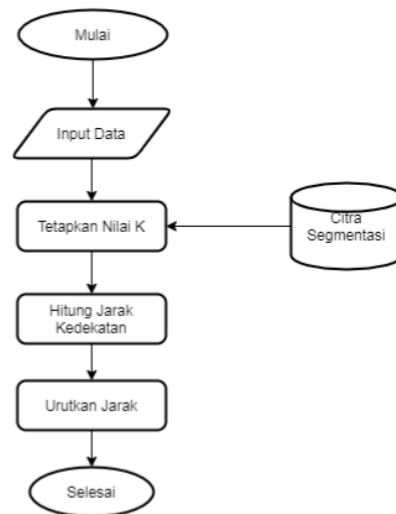
Parameter *eccentricity* merupakan teknik ekstraksi fitur yang bertujuan untuk mengambil atau mengekstraksi nilai *eccentricity*. Objek yang berbentuk memanjang atau mendekati bentuk garis lurus, nilai *eccentricity*-nya mendekati angka 1, sedangkan objek yang berbentuk bulat atau lingkaran, nilai

## C. Klasifikasi

Klasifikasi merupakan suatu proses pengkategorian yang dilakukan terhadap sekumpulan data. Klasifikasi diawali dengan membagi koleksi data menjadi dua bagian utama yaitu data latih dan data uji. Dari data latih di hitung dengan menggunakan metode tertentu kemudian, diperoleh model klasifikasi yang nantinya akan digunakan untuk penentuan kelas terhadap data uji[14].

## D. K-Nearest Neighbor (KNN)

Prinsip kerja K-Nearest Neighbor (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya dalam data pelatihan. Pengklasifikasian menggunakan fungsi jarak dari data baru ke data training (kedekatan). Ruang dibagi menjadi bagian-bagian berdasarkan klasifikasi data pelatihan. Algoritma diperlihatkan pada gambar 1[14].



Gambar 2. Algoritma K-Nearest Neighbor (KNN)

Keterangan Algoritma K-Nearest Neighbor:

1. Tentukan parameter K



2. Hitung jarak antara data yang akan dievaluasi dengan semua data latih
3. Urutkan jarak data
4. Tentukan jarak terdekat sampai urutan K
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang dievaluasi.

Rumus KNN ditunjukkan pada persamaan (1)

$$d_1 = \sqrt{\sum_{i=1}^p (X_1 - X_2)^2} \quad (1)$$

Keterangan:

$X_1$  = sampel data

$X_2$  = Data uji / testing

$i$  = Variabel

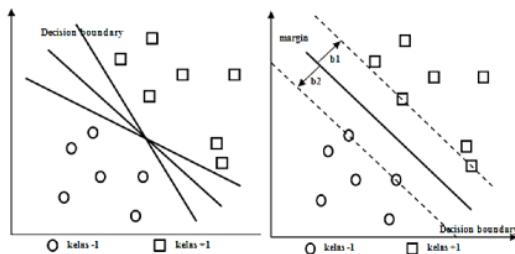
$d$  = Jarak

$p$  = Dimensi Data

Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari k objek. Algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari query instance yang baru. Algoritma metode KNN sangatlah sederhana, bekerja berdasarkan jarak terdekat dari query instance ke training sample untuk menentukan KNN-nya.

#### 4 E. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu teknik klasifikasi data dengan proses pelatihan (supervised learning). Support Vector Machine merupakan sebuah metode yang membandingkan suatu seleksi parameter standart nilai diskrit yang disebut kandidat set[15]. Pada dasarnya metode SVM bekerja dengan mendefinisikan batas antara dua kelas dengan jarak maksimal dari data yang terdekat[16]. Garis pemisah (hyperplane) yang terbaik merupakan salah satu ciri dari metode klasifikasi SVM untuk memperoleh ukuran margin yang maksimal. Untuk memaksimalkan batas hyperplane dijelaskan pada gambar 3.



Gambar 3. Batas Keputusan yang mungkin untuk set data

Menurut penjelasan [1] dari Konsep klasifikasi dengan SVM adalah sebagai cara dari mencari hyperplane terbaik yang berfungsi untuk pemisah dua buah kelas data input space. Pada gambar 2 memperlihatkan beberapa data yang merupakan anggota dari dua buah kelas data yakni +1 dan -1.

Sedangkan bentuk lingkaran pada gambar 2 merupakan symbol data yang tergabung dalam kelas -1, dan untuk kotak persegi merupakan symbol data yang tergabung dalam kelas +1. Hyperplane merupakan pemisah terbaik antara kedua kelas tersebut dapat ditemukan dengan mengukur margin hyperplane dan mencari titik maksimalnya. Margin merupakan jarak antara hyperplane tersebut dengan data terdekat dari masing-masing kelas. Data yang paling dekat inilah yang disebut dengan support vector[4]. Inti dari proses pelatihan pada SVM adalah usaha untuk mencari lokasi hyperplane.

Sequential Minimal Optimization (SMO) adalah algoritma untuk proses pelatihan SVM yang dapat memberikan solusi pada masalah optimisasi. Pada dasarnya penggunaan SVM hanya terbatas pada masalah yang kecil karena algoritma pelatihan SVM cenderung lambat, kompleks, dan sulit untuk diimplementasikan. Algoritma SMO akan memberikan bobot pada setiap atribut yang akan digunakan dalam klasifikasi[17].

#### F. Confusion Matrix

Metode ini menggunakan tabel matriks yang digunakan untuk membandingkan jumlah TP.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = \left( \frac{2}{Recall^{-1} + Precision^{-1}} \right) \quad (8)$$

Sebuah tabel klasifikasi binary untuk menentukan data yang digunakan bersifat True Positives (TP), False Positives (FP), False Negative (FN), dan True Negatives (TN). Seperti terlihat pada tabel 1[14].

TABEL I  
KLASIFIKASI BINARY

No	Klasifikasi yang Benar	Diklasifikasikan Sebagai	
		+	-
1	+	True Positives (TP)	False Negatives (FN)
2	-	False Positives (FP)	True Negatives (TN)

Keterangan:

TP = Jumlah true positives  
 TN = Jumlah true negatives  
 FN = Jumlah false negatives  
 FP = Jumlah false positives  
 P = Jumlah record positif  
 N = Jumlah tupel negatif

#### G. Tahapan Evaluasi

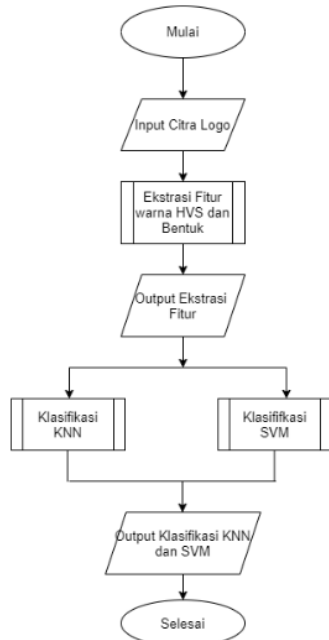
Penelitian ini menggunakan 10 gambar logo berformat png yang terdiri dari 1 gambar logo asli dan 9 gambar logo temipanulasi. Tujuan dari penelitian ini untuk membuktikan

berapa presentase keberhasilan yang lebih baik antara metode klasifikasi K-Nearest Neighbor (KNN) dan metode Support Vector Machine (SVM) dengan optimisasi SMO (Sequential Minimal Opimization). Untuk menguji tingkat keberhasilan peneliti menggunakan perhitungan akurasi confusion matrik. Untuk menghitung accuracy digunakan persamaan (6).

$$Accuracy = \frac{(a + c)}{(a + b + c + d)} \times 100\% \quad (9)$$

### III. METODE PENELITIAN

Metode yang digunakan pada penelitian ini ditunjukkan dengan langkah-langkah seperti diagram alir sistem pada gambar 4. Data yang digunakan adalah data primer yaitu data *image logo* yang buat menggunakan tools editing disini peneliti menggunakan *coreldraw* untuk menggambar logo sebagai data set dalam penelitian ini.



Gambar 4. Diagram Alir Penelitian

Tahapan penelitian meliputi, pertama pembuatan *image logo* menggunakan aplikasi *coreldraw X4* sebanyak data set yang akan dilatih dan diuji. Kedua input ke program untuk proses ekstraksi warna HSV dan bentuk. Ketiga hasil output ekstraksi fitur warna HSV dan bentuk dideskripsikan ke tabel dengan menggunakan format file .csv. Keempat file csv diubah menjadi file arff agar dapat diklasifikasi dengan dua metode yaitu *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)* dengan menggunakan WEKA.

#### A. Data Penelitian

Data penelitian diambil berupa *image logo* untuk evaluasi dalam *image matching*. Dalam penelitian ini kelas dan atribut dalam klasifikasi dibagi menjadi 2 kelas, yaitu bernilai 1 (citra asli) dan 0 (citra termanipulasi). Data keseluruhan berjumlah 10.

#### B. Ekstraksi Fitur Warna dan Bentuk

Ekstraksi Fitur warna HSV merupakan proses pemisahan objek dengan seleksi warna berdasarkan nilai Hue, Saturation, dan Value[18]. Ekstraksi Bentuk adalah proses mengukur nilai pada bentuk objek yaitu luas area, perimeter atau tepi area, dan eccentricity. Berikut pada gambar 5 menunjukkan alur proses ekstraksi warna dan bentuk. Parameter ekstraksi fitur warna dan bentuk didapat berdasarkan nilai yang ada pada objek tersebut.



Gambar 5. Proses Ekstraksi Warna dan Bentuk

Perubahan Ekstraksi Warna pada *image logo* seperti gambar 6. Nilai ekstraksi fitur dari Ekstraksi warna HSV dan bentuk dapat dilihat pada tabel 2. Setelah proses ekstraksi fitur warna dan bentuk *image logo*, selanjutnya penelitian ini akan mengklasifikasi *image* menggunakan data hasil segmentasi, menerapkan metode algoritma KNN dan SMO.



Gambar 6. Perubahan Segmentasi Citra

#### C. Klasifikasi

Klasifikasi yaitu menggunakan *Tools WEKA* serta membandingkan metode klasifikasi.

- 1) Menggunakan Klasifikasi K-Nearest Neighbor (KNN)











- 2) Support Vector Machine (SVM) dengan algoritma optimisasi SMO dalam tahap klasifikasi citra.

### III. HASIL DAN PEMBAHASAN

#### A. Tahapan Ekstraksi Ciri

Proses ekstraksi ciri yaitu proses mempartisi citra menjadi beberapa daerah atau objek. Dapat dilihat pada table 3, hasil segmentasi dataset yang akan diklasifikasi.

TABEL III  
HASIL EKSTRAKSI FITUR WARNA DAN BENTUK

Class Image	Hasil Ekstraksi Fitur Warna dan Bentuk					
	Hue	Saturation	Value	Area	Perimeter	Eccentricity
 Logo Asli	0.247	0.461	0.654	6641	1532.6	0.579
 Logo Termanipulasi	0.763	0.438	0.8453	8834	1051.13	0.273
 Logo Termanipulasi	0.277	0.727	0.905	1584	1059.8	0.865
 Logo Termanipulasi	0.040	0.829	0.9359	10416	523.1	0.860
 Logo Termanipulasi	0.139	0.7567	0.8768	23055	623.7	0.762
 Logo Termanipulasi	0.179	0.6278	0.9563	22453	847.7	0.900
 Logo Termanipulasi	0.494	0.712	0.938	2322	882.7	0.864
 Logo Termanipulasi	0.136	0.799	0.779	20459	1601.1	0.273
 Logo Termanipulasi	0.489	0.798	0.924	2284	861.6	0.890
 Logo Termanipulasi	0.737	0.672	0.880	1314	433.6	0.081

Logo Termanipulasi						
--------------------	--	--	--	--	--	--

#### B. Klasifikasi

Hasil dari uji data dalam pengujian sistem akan dilakukan menggunakan aplikasi Weka dengan menerapkan 10-fold *Cross-validation*. Hasil penilaian kinerja bersumber dari hasil pengukuran nilai *precision*, *recall*, *F-Measure* dan akurasi. Pada tahap pra-pengolahan data untuk menghilangkan anomali pada data file .cvs diubah menjadi .arff sebelum melakukan proses klasifikasi dengan menggunakan *tools* Weka.

##### 1) Menggunakan Klasifikasi K-Nearest Neighbor (KNN)

Metode Klasifikasi KNN menerapkan 10-fold Cross Validation dan Nilai K= 1, hasil *classification* data didapat seperti gambar 7 sebagai berikut:

```

Classifier output

10 Instance-based classifier
using 1 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8           80      %
Incorrectly Classified Instances    2           20      %
Kappa statistic                    -0.1111
Mean absolute error                 0.2064
Root mean squared error             0.4259
Relative absolute error             81.0829 %
Root relative squared error         127.0496 %
Total Number of Instances          10

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
0         0      0.111      0.000      0.000      0.000      0.889  LogoAsli
0.889      1      0.889      0.889      0.889      0.889      0.889  LogoMrip
Weighted Avg.  0.8      0.911      0.8      0.8      0.8      0.889

=== Confusion Matrix ===

 a b <-- classified as
 0 1 | a = LogoAsli
 1 8 | b = LogoMrip

```

Gambar 7. Hasil Klasifikasi KNN

Pada hasil klasifikasi KNN menunjukkan *correctly classified* sebanyak 80% dan *incorrectly classified* sebanyak 20%. Waktu untuk klasifikasi KNN 0 detik. Ini menunjukkan hasil klasifikasi KNN cukup baik dalam mengklasifikasi citra logo. Tabel 3 menunjukkan hasil akurasi dalam mengklasifikasi KNN.

TABEL III  
HASIL AKURASI KLASIFIKASI K-NEAREST NEIGHBOR (KNN)

	TP Rate	FP Rate	Precision	Recall	F-Meas	Roc Area	Classes
	0	0.11	0	0	0	0.889	Asli
	0.889	1	0.8	0.889	0.889	0.889	Mirip
W. Avg	0.8	0.911	0.8	0.8	0.8	0.889	

Dari hasil pengujian metode klasifikasi KNN menunjukkan hasil akurasi 80% dan  $\text{distanceWeighting} = \text{Weight by } 1 /$

distance, atau menerapkan nilai  $K = 1$  yaitu mencari nilai  $k$ -tetangga terdekat. Dari hasil tersebut dapat disimpulkan nilai  $K = 1$  memiliki nilai akurasi tertinggi yaitu 0.889.

## 2) Support Vector Machine (SVM)

Hasil klasifikasi menggunakan *Support Vector Machine* (SVM) dengan optimisasi SMO pada Tools Weka.

```

Classifier output
Number of kernel evaluations: 42 (75.700% cached)

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      7      70 %
Incorrectly Classified Instances    3      30 %
Kappa statistic                    -0.1538
Mean absolute error                  0.3
Root mean squared error             0.5477
Relative absolute error             117.8571 %
Root relative squared error         163.3743 %
Total Number of Instances          10

=== Detailed Accuracy By Class ===

      TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area   Class
0         0       0.222         0         0         0         0.389   LogoKali
1       0.778       1         0.875       0.778       0.824       0.389   LogoKali
Weighted Avg.   0.7       0.922       0.788       0.7       0.741       0.389

=== Confusion Matrix ===
a b <-- classified as
0 1 | a = LogoKali
2 7 | b = LogoKali

```

Gambar 8. Hasil Klasifikasi SVM dengan Optimisasi SMO

Dari hasil pengujian klasifikasi dengan SVM menggunakan optimisasi SMO menunjukkan Correctly Classified 70% dan Incorrectly Classified sebesar 30%. Dengan efisiensi waktu yang dibutuhkan untuk perhitungan metode ini 0.01 detik. Hasil uji menyatakan bahwa nilai akurasi terhadap metode klasifikasi KNN sebesar 80% dan klasifikasi metode SVM dengan optimisasi SMO sebesar 70%. Persentase dalam akurasi terbilang baik. Metode klasifikasi KNN lebih tinggi nilai akurasinya dan tidak membutuhkan waktu yang lama dalam *Classifier* dibandingkan optimisasi SMO.

Komparasi dari kedua metode ini menunjukkan hasil yang tidak terlalu jauh jika dibandingkan maka Metode klasifikasi *K-Nearest Neighbor* (KNN) lebih baik dalam mengklasifikasi citra logo asli dan logo termanipulasi daripada Metode klasifikasi Support Vector Machine (SVM) dengan menggunakan *Sequential Minimal Optimization* (SMO). Berdasarkan hasil komparasi dari kedua metode tersebut maka dapat disimpulkan bahwa untuk mengetahui tingkat kesamaan pada citra cukup melihat tingginya nilai akurasi pada penelitian, semakin tinggi maka nilai tersebut menunjukkan kemiripan objek sehingga dapat dinyatakan tingkat *plagiarism* tinggi, dan Metode *K-Nearest Neighbor* (KNN) menjadi metode yang cukup akurat dalam mengukur kemiripan citra.

## V. KESIMPULAN

Penelitian ini ingin mengukur presentase pencocokan citra (*image matching*) pada image logo dengan komparasi dari

kedua metode yang lebih baik antara metode klasifikasi KNN dan SVM optimisasi SMO dengan menggunakan *tools* Weka. Berdasarkan hasil pengujian pada komparasi *image matching* atau pencocokan citra dengan menggunakan klasifikasi *K-Nearest Neighbor* (KNN) dan Klasifikasi *Support Vector Machine* (SVM) dengan optimisasi SMO akurasi yang sangat tinggi adalah menggunakan Klasifikasi KNN dengan nilai akurasi sebesar 80% dan nilai kedekatan atau  $K$  antara citra yaitu 0,889 cukup baik untuk kedekatan dalam pencocokan citra menggunakan metode klasifikasi KNN, sedangkan dengan menggunakan SVM optimisasi SMO hasil akurasinya sebesar 70%. Maka untuk penelitian selanjutnya dapat mengembangkan metode dan memaksimalkan proses ekstraksi citra.

## DAFTAR PUSTAKA

- [1] T. T. Li, B. Jiang, Z. Z. Tu, B. Luo, and J. Tang, "Image matching using mutual k-nearest neighbor graph," *Commun. Comput. Inf. Sci.*, vol. 503, pp. 276–283, 2015.
- [2] J. I. Pengetahuan and D. A. N. T. Komputer, "Analisis Performa Algoritma Naive Bayes Pada," vol. 5, no. 1, pp. 37–42, 2019.
- [3] A. Ciputra, "D. 2. dan Algoritma Naive Bayes Dan Ekstraksi Fitur Citra Digital," vol. 9, no. 1, pp. 465–472, 2018.
- [4] K. Adi and R. R. Isnanto, "Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Teks Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM)," vol. 01, pp. 1–10, 2016, doi: 10.21456/vol6iss1pp1-10.
- [5] M. F. Grace and H. S. Scott, "An optional federal charter for insurance: Rationale and design," *Futur. Insur. Regul. United States*, vol. 6, no. 2, pp. 55–96, 2009.
- [6] & P. Rusydi Umar, Imam Riadi, "Perbandingan Metode SVM, RF dan SGD untuk Penentuan Model Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial," *J. Resti*, vol. 4, no. 2, pp. 329–335, 2020, [Online]. Available: <https://doi.org/10.29207/resti.v4i2.1770>.
- [7] K. Ayuningsih, Y. A. Sari, and P. P. Adikara, "Klasifikasi Citra Makanan Menggunakan HSV Color Moment dan Local Binary Pattern dengan Naive Bayes Classifier," vol. 3, no. 4, pp. 3166–3173, 2019.
- [8] R. Munir, "Pengantar Pengolahan Citra," *Pengolah. Citra Digit.*, no. Bagian 1, pp. 1–10, 2013.
- [9] P. N. Andono, *Pengolahan Citra Digital*. Yogyakarta: ANDI, 2015.
- [10] S. Gustina et al., "Identifikasi Tanaman Kamboja menggunakan Ekstraksi Ciri Citra Daun dan Jaringan Syaraf Tiruan," vol. 2, no. 1, pp. 128–132, 2016, [Online]. Available: <http://ars.ilkom.unsri.ac.id>.
- [11] E. Ciri, M. Gray, L. C. Matrix, I. Sel, and D. Putih, "JOINTECS Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix Untuk," vol. 3, no. 28, pp. 71–80, 2020.
- [12] G. A. Pradipta and P. D. Wulaning Ayu, "Perbandingan Segmentasi Citra Telur Ayam Menggunakan Metode Otsu Berdasarkan Perbedaan Ruang Warna Rgb Dan Hsv," *JST (Jurnal Sains dan Teknol.)*, vol. 6, no. 1, pp. 136–147, 2017, doi: 10.23887/jst-undiksha.v6i1.9329.
- [13] N. Arifin and I. S. Areni, "Klasifikasi Kematangan Stroberi Berbasis Segmentasi Warna dengan Metode HSV," vol. 23, no. 2, pp. 113–116, 2019, doi: 10.25042/jpe.112019.03.
- [14] E. F. Saraswita, "Akurasi Klasifikasi Citra Digital Scenes RGB Menggunakan Model K-Nearest Neighbor dan Naive Bayes," *Pros. Anu. Res. Semin.*, vol. 5, no. 1, pp. 978–979, 2019.
- [15] I. Riadi, R. Umar, and F. D. Aini, "Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm)," *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom.v11i1.361.17-24.
- [16] R. A. Rizal, I. S. Girsang, and S. A. Prasetyo, "Khsifikasi Wajah Menggunakan Support Vector Machine (SVM)," *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 3, no. 2, p. 1, 2019, doi: 10.33395/remik.v3i2.10080.



- [17] A. Wibowo, "Aplikasi Diagnosis Penyakit Kanker Payudara Menggunakan Algoritma Sequential Minimal Optimization," *J. Teknol. dan Sist. Komput.*, vol. 5, no. 4, p. 153, 2017, doi: 10.14710/jtsiskom.5.4.2017.153-158.
- [18] A. F. Hastawan, R. Septiana, and Y. E. Windarto, "Perbaikan Hasil Segmentasi Hsv Pada Citra Digital Menggunakan Metode Segmentasi Rgb Grayscale," *Edu Komputika J.*, vol. 6, no. 1, pp. 32–37, 2019, doi: 10.15294/EDUKOMPUTIKA.V6I1.23025.

# HASIL CEK\_60020397\_C73-Komparasi Image Matching Menggunakan Metode K-Nearest Neighbor (KNN) dan Metode Support Vector Machine (SVM)

## ORIGINALITY REPORT

10%

SIMILARITY INDEX

9%

INTERNET SOURCES

3%

PUBLICATIONS

4%

STUDENT PAPERS

## PRIMARY SOURCES

1

[jurnal.iaii.or.id](http://jurnal.iaii.or.id)

Internet Source

2%

2

[jurnal.fikom.umi.ac.id](http://jurnal.fikom.umi.ac.id)

Internet Source

2%

3

[repository.uin-suska.ac.id](http://repository.uin-suska.ac.id)

Internet Source

1%

4

[jurnal.atmaluhur.ac.id](http://jurnal.atmaluhur.ac.id)

Internet Source

1%

5

Submitted to Universitas Brawijaya

Student Paper

1%

6

[jurnal.umk.ac.id](http://jurnal.umk.ac.id)

Internet Source

1%

7

Submitted to Universitas Pamulang

Student Paper

1%

8

[filkom.ub.ac.id](http://filkom.ub.ac.id)

Internet Source

1%



---

Exclude quotes      On

Exclude bibliography      On

Exclude matches      < 1%